Sensitivity Analysis in Gaussian Bayesian Networks Using a Symbolic-Numerical Technique

Enrique Castillo[1] **and Uffe Kjærulff**[2]

[1] *Department of Applied Mathematics and Computational Sciences, University of Cantabria, Avda. Castros s/n, 39005 Santander, Spain.* (castie@ccaix3.unican.es)

[2] *Department of Computer Science, Aalborg University, Fredrik Bajers Vej 7, DK-9220, Aalborg, Denmark.* (uk@cs.aau.dk)

Abstract

The paper discusses the problem of sensitivity analysis in Gaussian Bayesian networks. The algebraic structure of the conditional means and variances, as rational functions involving linear and quadratic functions of the parameters, are used to simplify the sensitivity analysis. In particular the probabilities of conditional variables exceeding given values and related probabilities are analyzed. Two examples of application are used to illustrate all the concepts and methods.

Key Words: Sensitivity, Gaussian models, Bayesian networks.

1 Introduction

Sensitivity analysis is becoming an important and popular area of work. When solving practical problems, applied scientists are not satisfied enough with getting results coming from models, but they require a sensitivity analysis, indicating how sensitive the resulting numbers are to changes in the parameter values, to be performed (see Castillo, Gutiérrez and Hadi [4], Castillo, Gutiérrez, Hadi and Solares [5], Castillo, Solares, and Gómez.[6, 7, 8]).

In some cases, the parameter selection has an extreme importance in the final results. For example, it is well known how sensitive are the distributional assumptions and parameter values to tail distributions (see Galambos [14] or Castillo [2]). If this influence is neglected, the consequences can be disastrous. Thus, the relevance of sensitivity analysis.

Laskey [17] seems to be the first to address the complexity of sensitivity analysis of Bayesian networks, by introducing a method for computing the partial derivative of a posterior marginal probability with respect to a given parameter. Castillo, Gutiérrez and Hadi[5, 4] show that the function expressing the posterior probability is a quotient of linear functions in the parameters and the evidence values in the discrete case, and of the means, variances and evidence values, but covariances can appear squared. This discovery allows simplifying sensitivity analysis and making it computationally efficient (see, for example, Kjaerulff and van der Gaag [16], or Darwiche [12]).

In this paper we address the problem of sensitivity analysis in Gaussian Bayesian networks and show how changes in the parameter and evidence values influence marginal and conditional probabilities given the evidence.

This paper is structured as follows. In Section 2 we remind the reader about Gaussian Bayesian networks and introduce our working example. In Section 3 we discuss how to perform an exact propagation in Gaussian Bayesian networks. Section 4 is devoted to symbolic propagation. Section 5 analyses the sensitivity problem. Section 6 presents the damage of concrete structures example. Finally, Section 7 gives some conclusions.

2 Gaussian Bayesian Network Models

In this section we introduce Bayesian network models, but we first remind the reader the definition of Bayesian network.

Definition 1 (Bayesian network) *A Bayesian network is a pair $(\mathcal{G}, \mathcal{P})$, where $\mathcal{G}$ is a directed acyclic graph (DAG), $\mathcal{P} = \{p(x_1|\pi_1), \ldots, p(x_n|\pi_n)\}$ is a set of n conditional probability densities (CPD), one for each variable, and Π_i is the set of parents of node X_i in $\mathcal{G}$. The set P defines the associated joint probability density as*

$$p(\mathbf{x}) = \prod_{i=1}^{n} p(x_i|\pi_i). \tag{1}$$

The main two advantages of Bayesian networks are: (a) the factorization implied by (1), and (b) the fact that conditionally independence relations can be inferred directly from the graph $\mathcal{G}$.

Definition 2 (Gaussian Bayesian network) *A Bayesian network is said to be a Gaussian Bayesian network if and only if the JPD associated with its variables X is a multivariate normal distribution, $N(\mu, \Sigma)$, i.e., with joint probability density function:*

$$f(x) = (2\pi)^{-n/2}|\Sigma|^{-1/2} \exp\left\{-1/2(x-\mu)^T\Sigma^{-1}(x-\mu)\right\}, \tag{2}$$

where μ is the n-dimensional mean vector, Σ is the $n \times n$ covariance matrix, $|\Sigma|$ is the determinant of Σ, and μ^T denotes the transpose of μ.

Gaussian Bayesian networks have been treated, among others, by Kenley [15], Shachter and Kenley [22]), and Castillo, Gutiérrez and Hadi [3]. The JPD of the variables in a Gaussian Bayesian network can be specified as in (1) by the product of a set of CPDs whose joint probability density function is given by

$$f(x_i|\pi_i) \sim N\left(\mu_i + \sum_{j=1}^{i-1} \beta_{ij}(x_j - \mu_j),\ v_i\right), \tag{3}$$

where β_{ij} is the regression coefficient of X_j in the regression of X_i on the parents of X_i, Π_i, and

$$v_i = \Sigma_i - \Sigma_{i\Pi_i}\Sigma_{\Pi_i}^{-1}\Sigma_{i\Pi_i}^T$$

is the conditional variance of X_i, given $\Pi_i = \pi_i$, where Σ_i is the unconditional variance of X_i, $\Sigma_{i\Pi_i}$ is the covariances between X_i and the variables in Π_i, and Σ_{Π_i} is the covariance

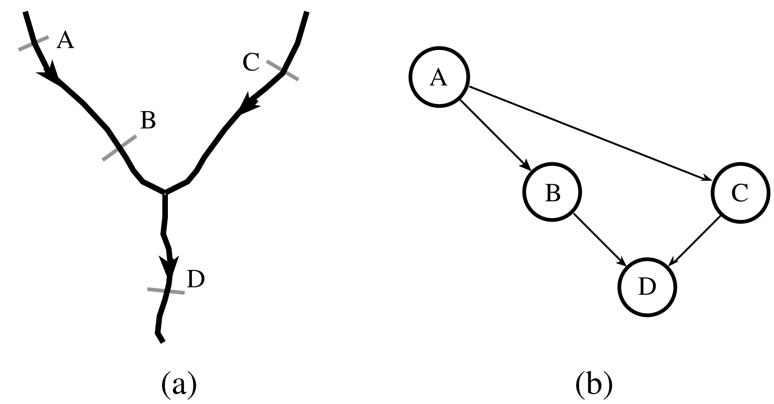

Figure 1: (a) The river in Example 1 and the selected cross sections, and (b) the Bayesian network used to solve the problem.

matrix of Π_i. Note that β_{ij} measures the strength of the relationship between X_i and X_j. If $\beta_{ij} = 0$, then X_j is not a parent of X_i.

Note that while the conditional mean $\mu_{x_i|\pi_i}$ depends on the values of the parents π_i, the conditional variance does not depend on these values. Thus, the natural set of CPDs defining a Gaussian Bayesian network is given by a collection of parameters $\{\mu_1, \ldots, \mu_n\}$, $\{v_1, \ldots, v_n\}$, and $\{\beta_{ij} \,|\, j < i\}$, as shown in (3).

The following is an illustrative example of a Gaussian Bayesian network.

Example 1 (Gaussian Bayesian network) Assume that we are studying the river in Figure 1(a), where we have indicated the four cross sections A, B, C and D, where the water discharges are measured. The mean time of the water going from A to B and from B to D is one day, and the mean time from C to D is two days. Thus, we register the set (A, B, C, D) with the corresponding delays. Assume that the joint water discharges can be assumed to be normal distributions and that we are interested in predicting B and D, one and two days later, respectively, from the observations of A and C. In Figure 1(b) we have shown the graph associated with a Bayesian network that shows the dependence structure of the variables involved.

Suppose that the random variable (A, B, C, D) is normally distributed, i.e., $\{A, B, C, D\} \sim N(\mu, \Sigma)$. A Gaussian Bayesian network is defined by specifying the set of CPDs appearing in the factorization (1), which gives

$$f(a, b, c, d) = f(a) f(b|a) f(c|a) f(d|b, c), \tag{4}$$

where

$$\begin{aligned}
&f(a) \sim N\left(\mu_A, v_A\right), \\
&f(b|a) \sim N\left(\mu_B + \beta_{BA}(a - \mu_A), v_B\right), \\
&f(c|a) \sim N\left(\mu_C + \beta_{CA}(a - \mu_A), v_C\right), \\
&f(d|b, c) \sim N\left(\mu_D + \beta_{DB}(b - \mu_B) + \beta_{DC}(c - \mu_C), v_D\right).
\end{aligned} \tag{5}$$

The parameters involved in this representation are $\{\mu_A, \mu_B, \mu_C, \mu_D\}$, $\{v_A, v_B, v_C, v_D\}$, and $\{\beta_{BA}, \beta_{CB}, \beta_{DB}, \beta_{DC}\}$.

Note that so far, all parameters have been considered in symbolic form. Thus, we can specify a Bayesian model by assigning numerical values to the parameters above. For example, for

$$\mu_A = 3, \;\; \mu_B = 4, \;\; \mu_C = 9, \;\; \mu_D = 14.$$

$$v_A = 4; \; v_B = 1; \; v_C = 4; \; v_D = 1, \; \beta_{BA} = 1, \; \beta_{CA} = 2, \; \beta_{DB} = 1, \; \beta_{DC} = 1.$$

we get

$$\mu = \begin{pmatrix} 3 \\ 4 \\ 9 \\ 14 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} 4 & 4 & 8 & 12 \\ 4 & 5 & 8 & 13 \\ 8 & 8 & 20 & 28 \\ 12 & 13 & 28 & 42 \end{pmatrix}.$$

■

3 Exact Propagation in Gaussian Networks

Several algorithms have been proposed in the literature to solve the problems of evidence propagation in these models. Some of them have originated from the methods for discrete models. For example, Normand and Tritchler [19] introduce an algorithm for evidence propagation in Gaussian network models using the same idea of the polytrees algorithm. Lauritzen [18] suggests a modification of the join tree algorithm to propagate evidence in mixed models.

Several algorithms use the structure provided by (1) and (3) for evidence propagation (see Xu and Pearl [23], and Chang and Fung [11]). In this section we present a conceptually simple and efficient algorithm that uses the covariance matrix representation. An incremental implementation of the algorithm allows updating probabilities, as soon as a single piece of evidence is observed. The main result is given in the following theorem, which characterizes the CPDs obtained from a Gaussian JPD (see, for example, Anderson [1]).

Theorem 1 Conditionals of a Gaussian distribution. *Let Y and Z be two sets of random variables having a joint multivariate Gaussian distribution with mean vector and covariance matrix given by*

$$\mu = \begin{pmatrix} \mu^Y \\ \mu^Z \end{pmatrix} \quad \text{and} \quad \Sigma = \begin{pmatrix} \Sigma^{YY} & \Sigma^{YZ} \\ \Sigma^{ZY} & \Sigma^{ZZ} \end{pmatrix},$$

where μ^Y and Σ^{YY} are the mean vector and covariance matrix of Y, μ^Z and Σ^{ZZ} are the mean vector and covariance matrix of Z, and $\Sigma^{YZ} = (\Sigma^{ZY})^T$ is the covariance of Y and Z. Then the CPD of Y given $Z = z$ is multivariate Gaussian with mean vector $\mu^{Y|Z=z}$ and covariance matrix $\Sigma^{Y|Z=z}$ that are given by

$$\mu^{Y|Z=z} = \mu^Y + \Sigma^{YZ} \Sigma^{ZZ^{-1}} (z - \mu^Z), \tag{6}$$

$$\Sigma^{Y|Z=z} = \Sigma^{YY} - \Sigma^{YZ} \Sigma^{ZZ^{-1}} \Sigma^{ZY}. \tag{7}$$

Note that the conditional mean $\mu^{Y|Z=z}$ depends on z but the conditional variance $\Sigma^{Y|Z=z}$ does not.

Theorem 1 suggests an obvious procedure to obtain the means and variances of any subset of variables $Y \subset X$, given a set of evidential nodes $E \subset X$ whose values are known to be $E = e$. Replacing Z in (6) and (7) by E, we obtain the mean vector and covariance matrix of the conditional distribution of the nodes in Y. Note that considering $Y = X \setminus E$ we get the joint distribution of the remaining nodes, and then we can answer questions involving the joint distribution of nodes instead of the usual information that refers only to individual nodes.

The methods mentioned above for evidence propagation in Gaussian Bayesian network models use the same idea, but perform local computations by taking advantage of the factorization of the JPD as a product of CPDs.

In order to simplify the computations, it is more convenient to use an incremental method, updating one evidential node at a time (taking elements one by one from E). In this case we do not need to calculate the inverse of a matrix because it degenerates to a scalar. Moreover, μ^Y and Σ^{YZ} are column vectors, and Σ^{ZZ} is also a scalar. Then the number of calculations needed to update the probability distribution of the nonevidential variables given a single piece of evidence is linear in the number of variables in X. Thus, this algorithm provides a simple and efficient method for evidence propagation in Gaussian Bayesian network models.

Due to the simplicity of this incremental algorithm, the implementation of this propagation method in the inference engine of an expert system is an easy task. The algorithm gives the CPD of the nonevidential nodes Y given the evidence $E = e$. The performance of this algorithm is illustrated in the following example.

Example 2 Propagation in Gaussian Bayesian network models. Consider the Gaussian Bayesian network given in Figure 1. Suppose we have the evidence $\{A = 7, C = 17, B = 8\}$.

If we apply expressions (6) and (7) to propagate evidence, we obtain the following:

After evidence $A = 7$: In the first iteration step, we consider the first evidential node $A = 7$. We obtain the following mean vector and covariance matrix for the rest of the nodes $Y = \{B, C, D\}$:

$$\mu^{Y|A=7} = \begin{pmatrix} 8 \\ 17 \\ 26 \end{pmatrix}; \; \Sigma^{YY|A=7} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 4 & 4 \\ 1 & 4 & 6 \end{pmatrix}. \tag{8}$$

After evidence $A = 7, C = 17$: The second step of the algorithm adds evidence $C = 17$; we obtain the following mean vector and covariance matrix for the rest of the nodes $Y = \{B, D\}$:

$$\mu^{Y|A=7,C=17} = \begin{pmatrix} 8 \\ 26 \end{pmatrix}; \; \Sigma^{YY|A=7,C=17} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}. \tag{9}$$

After evidence $A = 7, C = 17, B = 8$: Finally, after considering evidence $B = 8$ we get the conditional mean and variance of D, which are given by $\mu^{D|A=7,C=17,B=8} = 26$, $\Sigma^{DD|A=7,C=17,B=8} = 1$.

■

4 Symbolic Propagation in Gaussian Bayesian Networks

In Section 3 we presented several methods for exact propagation in Gaussian Bayesian networks. Some of these methods have been extended for symbolic computation (see, for example, Chang and Fung [11] and Lauritzen [18]). In this section we illustrate symbolic propagation in Gaussian Bayesian networks using the conceptually simple method given in Section 3. When dealing with symbolic computations, all the required operations must be performed by a program with symbolic manipulation capabilities unless the algebraic structure of the result be known. Figure 2 shows the *Mathematica* code for the symbolic implementation of the method given in Section 3. The code calculates the mean and variance of all nodes given the evidence in the evidence list.

Example 3 Consider the set of variables $X = \{A, B, C, D\}$ with mean vector and covariance matrix

$$\mu = \begin{pmatrix} p \\ 4 \\ 9 \\ q \end{pmatrix} \quad \text{and} \quad \Sigma = \begin{pmatrix} a & 4 & d & f \\ 4 & 5 & 8 & c \\ d & 8 & 20 & 28 \\ f & c & 28 & b \end{pmatrix}. \tag{10}$$

Note that some means and variances are specified in symbolic form, and that we have

$$\Sigma^{YY} = \begin{pmatrix} 5 & c \\ c & b \end{pmatrix}, \quad \Sigma^{ZZ} = \begin{pmatrix} a & d \\ d & 20 \end{pmatrix}, \quad \Sigma^{YZ} = \begin{pmatrix} 4 & 8 \\ f & 28 \end{pmatrix}. \tag{11}$$

We use the *Mathematica* code in Figure 2 to calculate the conditional means and variances of all nodes. The first part of the code defines the mean vector and covariance matrix of the Bayesian network. Table 1 shows the initial marginal probabilities of the nodes (no evidence) and the conditional probabilities of the nodes given each of the evidences $\{A = x_1\}$ and $\{A = x_1, C = x_3\}$. An examination of the results in Table 1 shows that the conditional means and variances are rational expressions, that is, ratios of polynomials in the parameters. Note, for example, that for the case of evidence $\{A = x_1, C = x_3\}$, the polynomials are first-degree in p, q, a, b, x_1, and x_3, that is, in the mean and variance parameters and in the evidence variables, and second-degree in d, f, i.e., the covariance parameters. Note also the common denominator for the rational functions giving the conditional means and variances. ■

The fact that the mean and variances of the conditional probability distributions of the nodes are rational functions of polynomials is given by the following theorem (see Castillo, Gutiérrez, Hadi, and Solares [5]).

Theorem 2 *Consider a Gaussian Bayesian network defined over a set of variables* $X = \{X_1, \ldots, X_n\}$ *with mean vector* μ *and covariance matrix* Σ*. Partition* X*,* μ*, and* Σ *as* $X = \{Y, Z\}$*,*

$$\mu = \begin{pmatrix} \mu^Y \\ \mu^Z \end{pmatrix}, \qquad \text{and} \qquad \Sigma = \begin{pmatrix} \Sigma^{YY} & \Sigma^{YZ} \\ \Sigma^{ZY} & \Sigma^{ZZ} \end{pmatrix},$$

```
(* Definition of the JPD *)
M={p,4,9,q};
V={{a, 4, d, f},
   {4, 5, 8, c},
   {d, 8, 20, 28},
   {f, c, 28, b}};
(* Nodes and evidence *)
X={A,B,C,D};
Ev={A,C};
ev={x1,x3};
(* Incremental updating of M and V *)
NewM=Transpose[List[M]];
NewV=V;
For[k=1, k<=Length[Ev], k++,
(* Position of the ith element of E[[k]] in X *)
    i=Position[X,Ev[[k]]][[1,1]];
    My=Delete[NewM,i];
    Mz=NewM[[i,1]];
    Vy=Transpose[Delete[Transpose[Delete[NewV,i]],i]];
    Vz=NewV[[i,i]];
    Vyz=Transpose[List[Delete[NewV[[i]],i]]];
    NewM=My+(1/Vz)*(ev[[k]]-Mz)*Vyz;
    NewV=Vy-(1/Vz)*Vyz.Transpose[Vyz];
(* Delete ith element *)
    X=Delete[X,i];
(* Printing results *)
    Print["Iteration step = ",k];
    Print["Remaining nodes = ",X];
    Print["M = ",Together[NewM]];
    Print["V = ",Together[NewV]];
    Print["--------------------"];
]
```

Figure 2: *Mathematica* code for symbolic propagation of evidence in a Gaussian Bayesian network model.

No Evidence		
Node	Mean	Variance
A	p	a
B	4	5
C	9	20
D	q	b
Evidence $A = x_1$		
Node	Mean $\mu^{Y\|A=x_1}$	Variance $\sigma^{YY\|A=x_1}$
A	x_1	0
B	$\frac{4(a-p+x_1)}{a}$	$\frac{5a-16}{a}$
C	$\frac{9a-dp+dx_1}{a}$	$\frac{20a-d^2}{a}$
D	$\frac{-fp+aq+fx_1}{a}$	$\frac{ab-f^2}{a}$
Evidence $A = x_1$ and $C = x_3$		
Node	Mean $\mu^{Y\|A=x_1,C=x_3}$	Variance $\sigma^{YY\|A=x_1,C=x_3}$
A	x_1	0
B	$\frac{4(2a+(9-d)d+(2d-20)p+(20-2d)x_1+(2a-d)x_3}{20a-d^2}$	$\frac{36a+64d-5d^2-320}{20a-d^2}$
C	x_3	0
D	$\frac{-252a+9df+(28d-20f)p+(20a-d^2)q}{20a-d^2} + \frac{(20f-28d)x_1+(28a-df)x_3}{20a-d^2}$	$\frac{(20ab-bd^2+56df-20f^2-784a}{20a-d^2}$

Table 1: Means and variances of the marginal probability distributions of nodes, initially and after evidence.

where μ^Y and Σ^{YY} are the mean vector and covariance matrix of Y, μ^Z and Σ^{ZZ} are the mean vector and covariance matrix of Z, and Σ^{YZ} is the covariance of Y and Z. Suppose that Z is the set of evidential nodes. Then the conditional probability distribution of any variable $X_i \in Y$ given Z is normal, with mean and variance that are ratios of polynomial functions in the evidential variables and the related parameters in μ and Σ. The polynomials involved are of degree at most one in the conditioning variables, in the mean, and in the variance parameters, and of degree at most two in the covariance parameters involving at least one Z (evidential) variable. Finally, the polynomial in the denominator is the same for all nodes.

In summary, we have

$$\mu^{Y|E=e} \quad = \quad \frac{\mu^Y \;\; |\Sigma^{EE}| + \Sigma^{YE} \;\; |adj\Sigma^{EE}| \;\; (e-\mu^E)}{|\Sigma^{EE}|}. \tag{12}$$

$$\Sigma^{Y|E=e} \quad = \quad \frac{\Sigma^{YY} \;\; |\Sigma^{EE}| - \Sigma^{YE} \;\; |adj\Sigma^{EE}| \;\; \Sigma^{EY}}{|\Sigma^{EE}|}. \tag{13}$$

Thus, we can conclude the following:

1. The parameters in μ^Y and μ^E appear in the numerator of the conditional means in linear form.

2. The parameters in Σ^{YY} appear in the numerator of the conditional variances in linear form.

3. The parameters in Σ^{YE} appear in the numerator of the conditional means and variances in linear and quadratic forms, respectively.

4. The variances and covariances in Σ^{EE} appear in the numerator and denominator of the conditional means and variances in linear, and linear or quadratic forms, respectively.

5. The evidence values appear only in the numerator of the conditional means in linear form.

Note that because the denominator polynomial is identical for all nodes, for implementation purposes it is more convenient to calculate and store all the numerator polynomials for each node and calculate and store the common denominator polynomial separately.

4.1 Extra Simplifications

Since the CDP $p(X_i = j|E = e)$ does not necessarily involve parameters associated with all nodes, we can identify the set of nodes which are relevant to the calculation of $p(X_i = j|E = e)$. Thus, important extra simplifications are obtained by considering only the set of parameters associated with the goal and the evidence variables. Doing this simplifications, dependences on all those parameters associated with the removed nodes, are known to be null.

Example 4 (The river example) *Assume that we are interested in calculating the probability $P(B > 11|A = 7, C = 17)$, that is, the target node is B and the evidential nodes A and C. Then, from (10), the mean and covariance matrix of $\{A, B, C\}$ are*

$$\mu = \begin{pmatrix} p \\ 4 \\ 9 \end{pmatrix} \text{ and } \Sigma = \begin{pmatrix} a & 4 & d \\ 4 & 5 & 8 \\ d & 8 & 20 \end{pmatrix}, \tag{14}$$

that implies independence on b, c, f and q.

Similarly, for the probability $P(D > 30|A = 7, C = 17)$, the mean and covariance matrix of $\{A, C, D\}$ are

$$\mu = \begin{pmatrix} p \\ 9 \\ q \end{pmatrix} \text{ and } \Sigma = \begin{pmatrix} a & d & f \\ d & 20 & 28 \\ f & 28 & b \end{pmatrix}. \tag{15}$$

that implies the independence on c.

5 Sensitivity Analysis

When dealing with Gaussian Bayesian networks, one is normally involved in calculating probabilities of the form:

$$\begin{array}{rcl} P(X_i > a|\mathbf{e}) & = & 1 - F_{X_i|\mathbf{e}}(a), \\ P(X_i \leq a|\mathbf{e}) & = & F_{X_i|\mathbf{e}}(a) \\ P(a < X_i \leq b|\mathbf{e}) & = & F_{X_i|\mathbf{e}}(b) - F_{X_i|\mathbf{e}}(a) \end{array} \tag{16}$$

and one is required to perform a sensitivity analysis on these probabilities with respect to a given parameter θ or evidence value e. Thus, it becomes important to know the partial derivatives

$$\frac{\partial F_{X_i|\mathbf{e}}(a; \mu(\boldsymbol{\theta}; \mathbf{e}), \sigma(\boldsymbol{\theta}; \mathbf{e}))}{\partial \theta} \text{ and } \frac{\partial F_{X_i|\mathbf{e}}(a; \mu(\boldsymbol{\theta}; \mathbf{e}), \sigma(\boldsymbol{\theta}; \mathbf{e}))}{\partial e}.$$

In what follows we use the compact notation $\boldsymbol{\rho} = (\boldsymbol{\theta}, \mathbf{e})$, and denote by ρ a single component of $\boldsymbol{\rho}$.

We can write

$$\frac{\partial F_{X_i|\mathbf{e}}(a; \mu(\boldsymbol{\rho}), \sigma(\boldsymbol{\rho}))}{\partial \rho} = \frac{\partial F_{X_i|\mathbf{e}}(a; \mu(\boldsymbol{\rho}), \sigma(\boldsymbol{\rho}))}{\partial \mu} \frac{\partial \mu(\boldsymbol{\rho})}{\partial \rho} + \frac{\partial F_{X_i|\mathbf{e}}(a; \mu(\boldsymbol{\rho}), \sigma(\boldsymbol{\rho}))}{\partial \sigma} \frac{\partial \sigma(\boldsymbol{\rho})}{\partial \rho}. \tag{17}$$

Since

$$F_{X_i|\mathbf{e}}(a; \mu(\boldsymbol{\rho}), \sigma(\boldsymbol{\rho})) = \Phi(\frac{a - \mu(\boldsymbol{\rho})}{\sigma(\boldsymbol{\rho})}) \tag{18}$$

we have

$$\frac{\partial F_{X_i|\mathbf{e}}(a; \mu(\boldsymbol{\rho}), \sigma(\boldsymbol{\rho}))}{\partial \mu} = f_{N(0,1)}(\frac{a - \mu(\boldsymbol{\rho})}{\sigma(\boldsymbol{\rho})}) \left(\frac{-1}{\sigma(\boldsymbol{\rho})} \right) \tag{19}$$

and

$$\frac{\partial F_{X_i|\mathbf{e}}(a; \mu(\boldsymbol{\rho}), \sigma(\boldsymbol{\rho}))}{\partial \sigma} = f_{N(0,1)}(\frac{a - \mu(\boldsymbol{\rho})}{\sigma(\boldsymbol{\rho})}) \left(\frac{\mu(\boldsymbol{\rho}) - a}{\sigma(\boldsymbol{\rho})^2} \right) \tag{20}$$

and then (17) becomes

$$\frac{\partial F_{X_i|\mathbf{e}}(a;\mu(\boldsymbol{\rho}),\sigma(\boldsymbol{\rho}))}{\partial\rho} = f_{N(0,1)}(\frac{a-\mu(\boldsymbol{\rho})}{\sigma(\boldsymbol{\rho})})\left[\left(\frac{-1}{\sigma(\boldsymbol{\rho})}\right)\frac{\partial\mu(\boldsymbol{\rho})}{\partial\rho} + \left(\frac{\mu(\boldsymbol{\rho})-a}{\sigma(\boldsymbol{\rho})^2}\right)\frac{\partial\sigma(\boldsymbol{\rho})}{\partial\rho}\right] \quad (21)$$

Thus, the partial derivatives $\frac{\partial F_{X_i|\mathbf{e}}(a;\mu(\boldsymbol{\rho}),\sigma(\boldsymbol{\rho}))}{\partial\rho}$ can be obtained by a single evaluation of $\mu(\boldsymbol{\rho})$ and $\sigma(\boldsymbol{\rho})$, and determining the partial derivatives $\frac{\partial\mu(\boldsymbol{\rho})}{\partial\rho}$ and $\frac{\partial\sigma(\boldsymbol{\rho})}{\partial\rho}$ with respect to all the parameters or evidence variables being considered. Thus, the calculus of these partial derivatives becomes crucial.

There are two ways of calculating these partial derivatives: (a) using the algebraic structure of the conditional means and variances, and (b) direct differentiations of the formulas (6) and (7). Here we use only the first method.

To calculate $\frac{\partial\mu_N(\boldsymbol{\rho})}{\partial\rho}$ and $\frac{\partial\sigma_N(\boldsymbol{\rho})}{\partial\rho}$ for node N we need to know the dependence of $\mu_N(\boldsymbol{\rho})$ and $\sigma_N(\boldsymbol{\rho})$ on the parameter or evidence variable ρ. This can be done with the help of Theorem 2. To illustrate, we use the previous example.

From Theorem 2 we can write

$$\mu_N^{Y|A=x_1,C=x_3}(a) = \frac{\alpha_1 a+\beta_1}{\gamma a+\delta}; \quad \sigma_N^{Y|A=x_1,C=x_3}(a) = \frac{\alpha_2 a+\beta_2}{\gamma a+\delta}, \quad (22)$$

where a is the parameter introduced in the equation (10), N is B or D, and since we have only 6 unknowns, calculation of $\mu_N^{Y|A=x_1,C=x_3}$ and $\sigma_N^{Y|A=x_1,C=x_3}$ for three different values of a allows determining the constant coefficients $\alpha_1,\alpha_2,\beta_1,\beta_2,\gamma$ and δ. Then, the partial derivatives with respect to a become

$$\frac{\partial\mu_N^{Y|A=x_1,C=x_3}(a)}{\partial a} = \frac{\alpha_1\delta-\beta_1\gamma}{(\gamma a+\delta)^2}; \quad \frac{\partial\sigma_N^{Y|A=x_1,C=x_3}(a)}{\partial a} = \frac{\alpha_2\delta-\beta_2\gamma}{(\gamma a+\delta)^2}. \quad (23)$$

Similarly, from Theorem 2 we can write

$$\mu_N^{Y|A=x_1,C=x_3}(f) = \frac{\alpha_3 f+\beta_3}{\gamma_1}; \quad \sigma_N^{Y|A=x_1,C=x_3}(f) = \frac{\gamma_4 f^2+\alpha_4 f+\beta_4}{\gamma_1} \quad (24)$$

where f is the parameter introduced in equation (10).

Since we have only 6 unknowns, calculation of a total of 6 values of $\mu_N^{Y|A=x_1,C=x_3}(f)$ and $\sigma_N^{Y|A=x_1,C=x_3}(f)$ for different values of f allows determining the constant coefficients $\alpha_3,\alpha_4,\beta_3,\beta_4$ and γ_1. Then, the partial derivatives with respect to f becomes

$$\frac{\partial\mu_N^{Y|A=x_1,C=x_3}(f)}{\partial f} = \frac{\alpha_3}{\gamma_1}; \quad \frac{\partial\sigma_N^{Y|A=x_1,C=x_3}(f)}{\partial f} = \frac{2\gamma_4 f+\alpha_4}{\gamma_1}. \quad (25)$$

It is worthwhile mentioning that if $N=B$, then $\alpha_3=\alpha_4=\beta_3=\beta_4=0$, and we need no calculations.

Finally, we can also obtain the partial derivatives with respect to evidence values. From Theorem 2 we can write

$$\mu_N^{Y|A=x_1,C=x_3}(x_1) = \alpha_5 x_1 + \beta_5; \quad \sigma_N^{Y|A=x_1,C=x_3}(x_1) = \gamma_2 \tag{26}$$

and since we have only 3 unknowns, calculation of a total of 3 values of $\mu_N^{Y|A=x_1,C=x_3}(x_1)$ and $\sigma_N^{Y|A=x_1,C=x_3}(x_1)$ for different values of x_1 allows determining the constant coefficients α_5, β_5 and γ_2. Then, the partial derivatives with respect to x_1 become

$$\frac{\partial \mu_N^{Y|A=x_1,C=x_3}(x_1)}{\partial x_1} = \alpha_5; \quad \frac{\partial \sigma_N^{Y|A=x_1,C=x_3}(x_1)}{\partial x_1} = 0. \tag{27}$$

It is worthwhile mentioning that if partial derivatives with respect to several parameters are to be calculated, the number of calculations reduces even more because some of them are common.

6 Damage of Concrete Structures

In this section we introduce a more complex example. The objective is to assess the damage of reinforced concrete structures of buildings. To this end, a Gaussian Bayesian network model is used.

The model formulation process usually starts with the selection of variables of interest. The goal variable (the damage of a reinforced concrete beam) is denoted by X_1. A civil engineer initially identifies 16 variables $(X_9, \ldots, X_{24})$ as the main variables influencing the damage of reinforced concrete structures. In addition, the engineer identifies seven intermediate unobservable variables $(X_2, \ldots, X_8)$ that define some partial states of the structure. Table 2 shows the list of variables and their definitions. The variables are measured using a scale that is directly related to the goal variable, that is, the higher the value of the variable the more the possibility for damage.

The next step in model formulation is the identification of the dependency structure among the selected variables. This identification is also given by a civil engineer.

In our example, the engineer specifies the following cause-effect relationships. The goal variable X_1, depends primarily on three factors: X_9, the weakness of the beam available in the form of a damage factor; X_{10}, the deflection of the beam; and X_2, its cracking state. The cracking state, X_2, is influenced in turn by four variables: X_3, the cracking state in the shear domain; X_6, the evaluation of the shrinkage cracking; X_4, the evaluation of the steel corrosion; and X_5, the cracking state in the flexure domain. Shrinkage cracking, X_6, depends on shrinkage, X_{23}, and the corrosion state, X_8. Steel corrosion, X_4, is influenced by X_8, X_{24}, and X_5. The cracking state in the shear domain, X_3, depends on four factors: X_{11}, the position of the worst shear crack; X_{12}, the breadth of the worst shear crack; X_{21}, the number of shear cracks; and X_8. The cracking state in the flexure domain, X_5 is affected by three variables: X_{13}, the position of the worst flexure crack; X_{22}, the number of flexure cracks; and X_7, the worst cracking state in the flexure domain. The variable X_{13} is influenced by X_4. The variable X_7 is a function of five variables: X_{14}, the breadth of the worst flexure

Table 2: Definitions of the variables related to the damage assessment example.

X_i	Definition
X_1	Damage assessment
X_2	Cracking state
X_3	Cracking state in shear domain
X_4	Steel corrosion
X_5	Cracking state in flexure domain
X_6	Shrinkage cracking
X_7	Worst cracking in flexure domain
X_8	Corrosion state
X_9	Weakness of the beam
X_{10}	Deflection of the beam
X_{11}	Position of the worst shear crack
X_{12}	Breadth of the worst shear crack
X_{13}	Position of the worst flexure crack
X_{14}	Breadth of the worst flexure crack
X_{15}	Length of the worst flexure cracks
X_{16}	Cover
X_{17}	Structure age
X_{18}	Humidity
X_{19}	PH value in the air
X_{20}	Content of chlorine in the air
X_{21}	Number of shear cracks
X_{22}	Number of flexure cracks
X_{23}	Shrinkage
X_{24}	Corrosion

crack; X_{15}, the length of the worst flexure crack; X_{16}, the cover; X_{17}, the structure age; and X_8, the corrosion state. The variable X_8 is affected by three variables: X_{18}, the humidity; X_{19}, the PH value in the air; and X_{20}, the content of chlorine in the air.

All these relationships are summarized in the directed acyclic graph of Figure 3, that is the network structure of the selected Gaussian Bayesian network model.

Then, the next step is the definition of the JPD. Suppose that the means of all variables are zeros, the coefficients β_{ij} in (3) are defined as shown in Figure 3, and the conditional variances are given by

$$v_i = \begin{cases} 10^{-4}, & \text{if } X_i \text{ is unobservable,} \\ 1, & \text{otherwise.} \end{cases}$$

To propagate evidence in the above Gaussian Bayesian network model, we use the incremental algorithm described in Theorem 1. For illustrative purpose, we assume that the engineer examines a given concrete beam and sequentially obtains the values $\{x_9, x_{10}, \ldots, x_{24}\}$

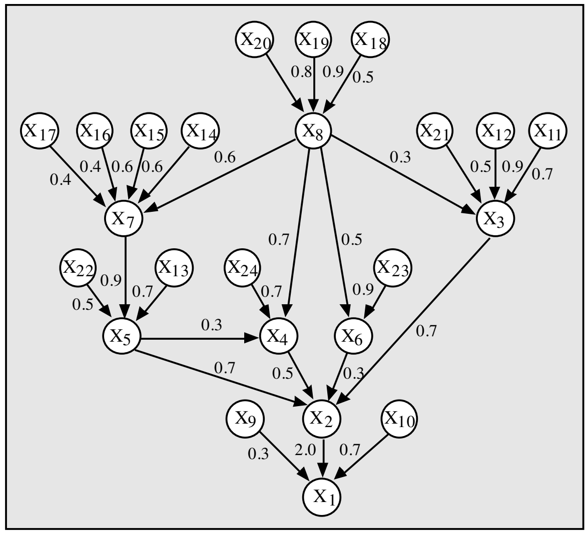

Figure 3: Directed graph for the damage assessment of reinforced concrete structure.

corresponding to the observable variables $X_9, X_{10}, \ldots, X_{24}$. For the sake of simplicity, suppose that the obtained evidence is $e = \{X_9 = 1, \ldots, X_{24} = 1\}$, which indicates serious damage of the beam.

Again, we wish to assess the damage (the goal variable, X_1). The conditional mean vector and covariance matrix of the remaining (unobservable and goal) variables $Y = (X_1, \ldots, X_8)$ given e, obtained using the incremental algorithm, are

$$E(y|e) = (2.2, 3.32, 2.0, 4.19, 3.50, 2.76, 7.21, 15.42),$$

$$Var(y|e) = \begin{pmatrix} 0.00010 & \ldots & 0.00009 & 0.00003 & 0.00012 & 0.00023 \\ 0.00006 & \ldots & 0.00008 & 0.00002 & 0.00015 & 0.00029 \\ 0.00005 & \ldots & 0.00004 & 0.00001 & 0.00009 & 0.00018 \\ 0.00005 & \ldots & 0.00010 & 0.00002 & 0.00022 & 0.00043 \\ 0.00009 & \ldots & 0.00019 & 0.00003 & 0.00020 & 0.00039 \\ 0.00003 & \ldots & 0.00003 & 0.00011 & 0.00011 & 0.00021 \\ 0.00012 & \ldots & 0.00020 & 0.00010 & 0.00045 & 0.00090 \\ 0.00023 & \ldots & 0.00039 & 0.00021 & 0.00090 & 1.00200 \end{pmatrix}.$$

Thus, the conditional distribution of the variables in Y is normal with the above mean vector and variance matrix.

Note that in this case, all elements in the covariance matrix except for the conditional variance of X_1 are close to zero, indicating that the mean values are very good estimates for $E(X_2, \ldots, X_8)$ and a reasonable estimate for $E(X_1)$.

We can also consider the evidence sequentially. Table 3 shows the conditional mean and variance of X_1 given that the evidence is obtained sequentially in the order given in the table. The evidence ranges from no information at all to complete knowledge of all the observed values $x_9, x_{10}, \ldots, x_{24}$. Thus, for example, the initial mean and variance of X_1 are $E(X_1) = 0$ and $Var(X_1) = 19.26$, respectively; and the conditional mean and variance

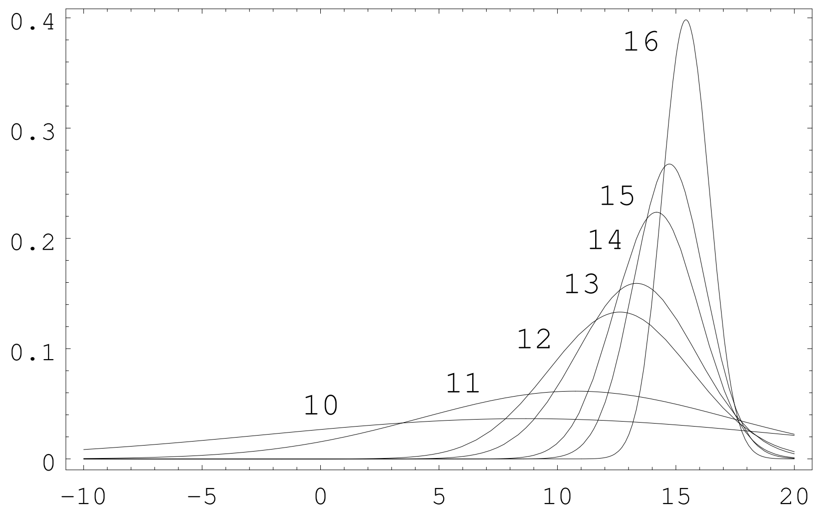

Figure 4: Conditional distributions of node X_1 corresponding to the cumulative evidence in Table 3. The step number is written near the curve.

of X_1 given $X_9 = 1$ are $E(X_1|X_9 = 1) = 0.30$ and $Var(X_1|X_9 = 1) = 19.18$. Note that after observing the key evidence $X_9 = 1$, the mean of X_1 has increased from 0 to 0.3 and the variance has decreased from 19.26 to 19.18. As can be seen in the last row of the table, when all the evidences are considered, $E(X_1|X_9 = 1, \ldots, X_{24} = 1) = 15.42$ and $Var(X_1|X_9 = 1, \ldots, X_{24} = 1) = 1.0$, an indication that the building is seriously damaged. In Figure 4 we show several of the conditional normal distributions of X_1 when a new evidence is added. The figure shows the increasing damage of the beam at different steps, as would be expected. Note that the mean increases and the variance decreases in almost all cases, an indication of decreasing uncertainty.

Step	Available Evidence	Damage Mean	Damage Variance	Step	Available Evidence	Damage Mean	Damage Variance
0	None	0.00	19.26	9	$X_{17} = 1$	7.49	12.29
1	$X_9 = 1$	0.30	19.18	10	$X_{18} = 1$	8.70	10.92
2	$X_{10} = 1$	1.00	18.69	11	$X_{19} = 1$	10.76	6.49
3	$X_{11} = 1$	1.98	17.73	12	$X_{20} = 1$	12.63	2.99
4	$X_{12} = 1$	3.24	16.14	13	$X_{21} = 1$	13.33	2.51
5	$X_{13} = 1$	4.43	17.72	14	$X_{22} = 1$	14.18	1.78
6	$X_{14} = 1$	5.35	13.88	15	$X_{23} = 1$	14.72	1.49
7	$X_{15} = 1$	6.27	13.04	16	$X_{24} = 1$	15.42	1.00
8	$X_{16} = 1$	6.88	12.66				

Table 3: Mean and variances of the damage, X_1, given the cumulative evidence of $x_9, x_{10}, \ldots, x_{24}$.

Suppose now that we are interested in the effect of the deflection of the beam, X_{10}, on the goal variable, X_1. Then, we consider X_{10} as a symbolic node. Let $E(X_{10}) = m, Var(X_{10}) = v, Cov(X_{10}, X_1) = Cov(X_1, X_{10}) = c$. The conditional means and variances of all nodes are calculated by applying the algorithm for symbolic propagation in Gaussian Bayesian networks introduced in Figure 2. The conditional means and variances of X_1 given the sequential evidences $X_9 = 1, X_{10} = 1, X_{11} = x_{11}, X_{12} = 1, X_{13} = x_{13}, X_{14} = 1$, are shown in Table 4. Note that some of the evidences (X_{11}, X_{13}) are given in a symbolic form.

Note that the values in Table 3 are a special case of those in Table 4. They can be obtained by setting $m = 0$, $v = 1$, and $c = 0.7$ and considering the evidence values $X_{11} = 1, X_{13} = 1$. Thus the means and variances in Table 3 can actually be obtained from Table 4 by replacing the parameters by their values. For example, for the case of the evidence $X_9 = 1, X_{10} = 1, X_{11} = x_{11}$, the conditional mean of X_1 is $(c - cm + 0.3v + 0.98vx_{11})/v = 1.98$. Similarly, the conditional variance of X_1 is $(-c^2 + 18.22v)/v = 17.73$.

Available Evidence	Damage	
	Mean	Variance
None	0	19.26
$X_9 = 1$	0.3	19.18
$X_{10} = 1$	$\frac{c - cm + 0.3v}{v}$	$\frac{-c^2 + 19.18v}{v}$
$X_{11} = x_{11}$	$\frac{c - cm + 0.3v + 0.98vx_{11}}{v}$	$\frac{-c^2 + 18.22v}{v}$
$X_{12} = 1$	$\frac{c - cm + 1.56v + 0.98vx_{11}}{v}$	$\frac{-c^2 + 16.63v}{v}$
$X_{13} = x_{13}$	$\frac{c - cm + 1.56v + 0.98vx_{11} + 1.19vx_{13}}{v}$	$\frac{-c^2 + 15.21v}{v}$
$X_{14} = 1$	$\frac{c - cm + 2.48v + 0.98vx_{11} + 1.19vx_{13}}{v}$	$\frac{-c^2 + 14.37v}{v}$

Table 4: Conditional means and variances of X_1, initially and after cumulative evidence.

7 Conclusions

From the previous sections, the following conclusions can be obtained:

1. Sensitivity analysis in Gaussian Bayesian networks is greatedly simplified due to the knowledge of the algebraic structure of conditional means and variances.

2. The algebraic structure of any conditional mean or variance is a rational function of the parameters.

3. The degrees of the numerator and denominator polynomials in the parameters can be immediately identified, as soon as the parameter or evidence value is defined.

4. Closed expressions for the partial derivatives of probabilities of the form $P(X_i > a|\mathbf{e})$, $P(X_i \leq a|\mathbf{e})$ and $P(a < X_i \leq b|\mathbf{e})$ with respect to the parameters, or evidence values, can be obtained.

5. Much more that sensitivity measures can be obtained. In fact, closed formulas for the probabilities $P(X_i > a|\mathbf{e})$, $P(X_i \leq a|\mathbf{e})$ and $P(a < X_i \leq b|\mathbf{e})$ as a function of the desired parameter, or evidence value, can be obtained.

References

[1] Anderson, T. W. (1984), *An introduction to multivariate statistical analysis, 2nd edition.* John Wiley and Sons, New York.

[2] Castillo, E. *Extreme value theory in engineering.* Academic Press, New York, 1988.

[3] Castillo, E., Gutiérrez, J. M. and Hadi, A. S. *Expert systems and probabilistic network models.* Springer Verlag, New York, 1997.

[4] Castillo, E., Gutiérrez, J. M. and Hadi, A. S. Sensitivity analysis in discrete Bayesian networks. *IEEE Transactions on Systems, Man and Cybernetics*, Vol 26, N. 7, 412–423, 1997.

[5] Castillo, E., Gutiérrez, J. M., Hadi, A. S. and Solares, C. Symbolic propagation and sensitivity analysis in Gaussian Bayesian networks with application to damage assessment. *Artificial Intelligence in Engineering*, 11:173–181, 1997.

[6] Castillo, E., Solares, C., and Gómez, P. Tail sensitivity analysis in Bayesian networks. In Proceedings of the Twelfth Conference on Uncertainty in Artificial Intelligence (UAI'96), Portland (Oregon), Morgan Kaufmann Publishers, San Francisco, California, 133–140, 1996.

[7] Castillo, E., Solares, C., and Gómez, P. Estimating extreme probabilities using tail simulated data. *International Journal of Approximate Reasoning*, 1996.

[8] Castillo, E., Solares, C., and Gómez, P. High probability one-sided confidence intervals in reliability Models. *Nuclear Science and Engineering*, Vol. 126, 158–167, 1997.

[9] Castillo, E., Solares, C., and Gómez, P. Tail uncertainty analysis in complex systems. *Artificial Intelligence* 96(2), 395-419, 1997.

[10] Castillo, E., Sarabia, J. M., Solares, C., and Gómez, P. Uncertainty analyses in fault trees and Bayesian networks using FORM/SORM methods. *Reliability Engineering and System Safety (in press)*, 1998.

[11] Chang, K. C. and Fung, R. (1991), Symbolic probabilistic inference with continuous variables. In *Proceedings of the Seventh Conference on Uncertainty in Artificial Intelligence*. Morgan Kaufmann Publishers, San Mateo, CA, 77–85.

[12] Darwiche, A. A differential approach to inference in Bayesian networks. UAI2000.

[13] Geiger, D., Verma, T., and Pearl, J. 1990. Identifying Independence in Bayesian Networks. *Networks*, 20:507−534.

[14] Galambos, J. *The asymptotic theory of extreme order statistics*. Robert E. Krieger Publishing Company. Malabar, Florida, 1987.

[15] Kenley, C. R. (1986), *Influence diagram models with continuous variables*. Ph.D. Thesis, Stanford, Stanford University.

[16] Kjaerulff, U. and van der Gaag, L. C. Making sensitivity analysis computationally efficient. UAI2000.

[17] Laskey, K. B. Sensitivity analysis for probability assessments in Bayesian networks, *IEEE Transactions on Systems, Man, and Cybernetics*, **25**, 901–909.

[18] Lauritzen, S. L. (1992), Propagation of probabilities, means, and variances in mixed graphical association models. *Journal of the American Statistical Association*, 87:1098–1108.

[19] Normand, S.-L. and Tritchler, D. (1992), Parameter updating in Bayes network. *Journal of the American Statistical Association*, 87:1109–1115.

[20] M. Rosenblatt. Remarks on a multivariate transformation. *Ann. Math. Stat. 23(3)*,470–472, 1952.

[21] Shachter, R. D. 1990. An Ordered Examination of Influence Diagrams. *Networks*, 20:535−563.

[22] Shachter, R. and Kenley, C. (1989), Gaussian influence diagrams. *Management Science*, 35(5):527–550.

[23] Xu, L. and Pearl, J. (1989), Structuring causal tree models with continuous variables. In Kanal, L. N., Levitt, T. S., and Lemmer, J. F., editors, *Uncertainty in Artificial Intelligence 3*. North Holland, Amsterdam, 209–219.